HAL
open science

Using single visits into integrated occupancy models to make the most of existing monitoring programs

Valentin Lauret, Hélène Labach, Matthieu Authier, Olivier Gimenez

To cite this version:

Valentin Lauret, Hélène Labach, Matthieu Authier, Olivier Gimenez. Using single visits into integrated occupancy models to make the most of existing monitoring programs. Ecology, 2021, 102 (12), 10.1002/ecy.3535 . hal-03649791

HAL Id: hal-03649791

https://hal.science/hal-03649791v1

Submitted on 23 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Running head: Integrated single-visit occupancy models

Title: Using single visits into integrated occupancy models to make the most of existing monitoring programs

Authors: Valentin Lauret[1], Hélène Labach[1,2], Matthieu Authier[3,4], Olivier Gimenez[1]

(1) CEFE, Université Montpellier, CNRS, EPHE, IRD, Université Paul Valéry Montpellier 3, Montpellier, France

(2) MIRACETI, Connaissance et conservation des cétacés, Place des traceurs de pierres, 13500 La Couronne, France

(3) ADERA, 162 avenue Albert Schweitzer, 33608 Pessac Cedex

(4) Observatoire PELAGIS, UMS 3462 CNRS-La Rochelle Université, 5 allée de l'Océan, 17000 La Rochelle

Corresponding author: Valentin Lauret, valentin.lauret@ens-lyon.fr, CEFE 1919 Route de Mende, 34090 Montpellier, France.

Open Research Statement: This submission uses novel code, which is provided as Supporting Information to be evaluated as part of the submission.

After publication, codes and data will be stored on the following permanent repository https://github.com/valentinlauret/IntegratedSingleVisitOccupancy

Abstract

A major challenge in statistical ecology consists of integrating knowledge from different datasets to produce robust ecological indicators. To estimate species distribution, occupancy models are a flexible framework that can accommodate several datasets obtained from different sampling methods. However, repeating visits at sampling sites is a prerequisite for using standard occupancy models. Occupancy models were recently developed to analyze detection/non-detection data collected during a single visit. To date, single-visit occupancy models have never been used to integrate several different datasets. Here, we showcase an approach that combines two datasets into an integrated single-visit occupancy model. As a case study, we estimated the distribution of common bottlenose dolphin (*Tursiops truncatus*) over the North-western Mediterranean Sea by combining 24,624 km of aerial surveys and 21,464 km of at-sea monitoring. We compared the outputs of single- vs. repeated-visit occupancy models into integrated occupancy models. Integrated models allowed a better sampling coverage of the targeted population, which provided a better precision for occupancy estimates than occupancy models using datasets in isolation. Overall, single- and repeated-visit integrated occupancy models produced similar inference about the distribution of bottlenose dolphins. We suggest that single-visit occupancy models open promising perspectives for the use of existing ecological datasets.

Keywords

Bottlenose dolphins, Ecological monitoring, Integrated species distribution models, Multi-method, Occupancy models, Single-visit

Introduction

In large-scale ecological analysis, several parallel monitoring programs are often carried out to collect ecological data (Zipkin and Saunders 2018). Ecological monitoring programs are conducted by organizations operating across different time scales, geographic scales and

funding initiatives (Lindenmayer and Likens 2010). A major challenge is integrating knowledge from different monitoring programs to produce robust ecological indicators that may be used to inform decision-making (Fletcher et al. 2019, Zipkin et al. 2021). Recently, modelling tools have emerged to combine multiple data sources to estimate species distributions and *Integrated models* refer to the approaches that combine different data sources (Miller et al. 2019, Isaac et al. 2019). The main purpose of integrated models is to improve the accuracy of ecological indicators (Fletcher et al. 2019, Zipkin et al. 2019). Species distributed over large areas could particularly benefit from integrated models because they allow for a global coverage of species occurrence by combining different data sources collected at different spatial scales (Miller et al. 2019). To estimate species distribution in the face of uncertainties inherent to data collection, occupancy models are commonly used statistical tools (Mackenzie et al., 2002). Occupancy models have been developed to estimate species distribution while accounting for false negatives in the observation process (Mackenzie et al. 2002). Estimating occupancy when species detection is not perfect requires performing *repeated visits* to a set of sites to assess the detection probability (MacKenzie 2006). However, repeating visits is sometimes unfeasible due to associated costs and logistical issues. In this context, two relevant developments of occupancy models have been recently proposed. First, integrated occupancy models combine data from different monitoring programs to improve the estimation of species distribution (Nichols et al. 2008, Miller et al. 2019, Fletcher et al. 2019). Second, Lele et al., (2012) used occupancy models to estimate species distribution and detectability while having only one visit at the sampling site, i.e. hereafter *single-visit* occupancy models. An increasing number of studies suggest that under certain conditions, single-visit models produce robust estimates of occupancy without repeating visits at the sampling sites (Lele et al. 2012, Sólymos and Lele 2016, Peach et al. 2017). Besides, single-visit occupancy offers the possibility to work with existing datasets that did not carry out repeated visits, which is relevant to population

biology and management. In this paper, we develop an integrated approach that combines two single-visit occupancy models and demonstrate that combining several datasets into integrated single-visit occupancy models leads to accurate ecological parameter estimation. We also investigate the performance of single-visit vs. repeated-visit occupancy models. As a case study, we focused on the distribution of Bottlenose dolphins (*Tursiops truncatus*) in the North-Western Mediterranean Sea. We combined aerial surveys and at-sea monitoring into integrated occupancy models and we compared the outputs of integrated occupancy models to occupancy models using each dataset in isolation. Last, we discuss the advantages of integrated single-visit occupancy models to deal with existing ecological monitoring programs.

Methods

MODEL DESCRIPTION

Latent ecological process

Occupancy models estimate spatial distribution while accounting for imperfect species detection (Mackenzie et al. 2002). The formulation of occupancy models as state-space models allows distinguishing the latent ecological state process (i.e. species present or absent at a grid-cell) from the detection process (Royle and Kéry 2007). We denote z_i the latent occupancy of grid-cell i (z = 1, presence; z = 0, absence). We assume z_i is drawn from a Bernoulli distribution with Ψ_i the probability that the species is present at grid-cell i:

$$z_i \sim \text{Bernoulli}(\Psi_i)$$

We modelled Ψ as a function of some environmental covariate on a logit scale, say habitat. $\text{logit}(\Psi_i) = \beta_0 + \beta_1 \text{habitat}_i$ where parameters β_0, and β_1 are to be estimated.

Repeated-visit observation process

In standard occupancy designs, each grid-cell is visited J times to estimate the detection probability. We denote $y_{i,j}$ ($y_{i,j}$ = 0, no detection ; $y_{i,j}$ = 1, detection) the observations corresponding to the data collected at grid-cell i during visit j (j =1,..,J). Repeating visits at a

grid-cell allows estimating species detectability, with $p_{i,j}$ being the probability of detecting the species at visit j given it is present at grid-cell i:

$$y_{i,j} \mid z_i \sim \text{Bernoulli}(z_i\, p_{i,j})$$

Single-visit observation process

The difference with repeated-visit occupancy models lies in the number of sampling occasions which is $J = 1$ in single-visit occupancy models. The j subscript is dropped and we denoted y_i the observation corresponding to the data collected at site i. Subsequently, p_i is the probability of detecting the species during the single visit given it is present at site i: $y_i \mid z_i \sim \text{Bernoulli}(z_i\, p_i)$. Single-visit occupancy models require certain conditions to be fulfilled for estimating detection probabilities reliably. First, different continuous covariates should be used to estimate detection and occupancy probabilities (Lele et al. 2012, Peach et al. 2017). Second, the number of detections may affect the estimation of occupancy in the case of rare or ubiquitous species (Peach et al. 2017). Third, the use of inappropriate link functions to model the detection process may lead to model misspecification and biased interpretation (e.g. log-link and scaled logit link function on detection, Knape & Korner-Nievergelt, 2015). However, most often, the logit link function is used for detection, which makes the single-visit approach valid (Sólymos and Lele 2016). Despite these concerns, simulation studies have showed that situations where single-visit occupancy models fail are rare (Sólymos and Lele 2016, Peach et al. 2017) and, in practice, the conditions for a valid application of single-visit occupancy models are often fulfilled (Sólymos and Lele 2016). We detailed the modeling assumptions of single-visit occupancy models in Appendix S4. Because the number of detections is an important condition to accurately estimate single-visit occupancy parameters (Peach et al. 2017), we expect that integrated approaches will be beneficial to single-visit occupancy modelling by increasing the number of detections (true occupancy) available.

Integrated occupancy models

We developed an integrated occupancy model using data from two independent monitoring programs, say A and B. The state process driving the latent occupancy state of site i, z_i, remains unchanged and is drawn from a Bernoulli distribution with probability ψ, which is modeled as a function of environmental covariates. The observation of the targeted species at site i during occasion j may take four values with $y_{i,j} = 0$ for no detection, $y_{i,j} = 1$ for detection in dataset A, $y_{i,j} = 2$ for detection in dataset B, and $y_{i,j} = 3$ for detection in both datasets A and B. For convenience, we drop the subscripts in the notation as the formulation of the integrated observation process is identical whether we consider single-visit occupancy (i.e. $J = 1$) or repeated-visit occupancy ($J > 1$). Assuming that detection methods are independent, the observation process can be written using detection probability by the monitoring program A (p_A) and detection probability by the monitoring program B (p_B):

$$y|z \sim Multinomial(1, z\pi) \text{ with}$$

$$\pi = [p_0, p_1, p_2, p_3] = [pr(y = 0), pr(y = 1), pr(y = 2), pr(y = 3)]$$

$$\pi = [1 - p_A - p_B + p_A p_B, p_A(1 - p_B), p_B(1 - p_A), p_A p_B]$$

We modeled monitoring-specific detection probabilities as functions of the sampling effort of each monitoring program:

$$\text{logit}(p_A) = \alpha_{0A} + \alpha_{1A} \log(\text{Seff}_A)$$

$$\text{logit}(p_B) = \alpha_{0B} + \alpha_{1B} \log(\text{Seff}_B).$$

where the parameters α_{0A}, α_{1A}, α_{0B}, and α_{1B} are to be estimated. For example, if we assume that the detection history at site i is $y_i = \{2,0,1,2\}$ over $J = 4$ sampling occasions, i.e. the species was detected by monitoring program B only at sampling occasions $j = 1$ and $j = 4$, then went undetected at $j = 2$, and was detected by monitoring program A only at $j = 3$, then for single-visit integrated occupancy we consider $y_i = \{3\}$ because both monitoring programs detected the species at site i. We ran a simulation study comparing the performance of single- vs. repeated-

144 visit occupancy over different scenarios affecting occupancy, and detection probabilities
145 (Appendix S1).

146 BOTTLENOSE DOLPHINS CASE STUDY

147 We aimed at estimating bottlenose dolphin (*Tursiops truncatus*) distribution in an area
148 of 255,000 km^2 covering the North-Western Mediterranean. The protected status of this species
149 within the French seas led to the development of specific programs to monitor Mediterranean
150 bottlenose dolphins within the implementation of the European Marine Strategy Framework
151 Directive (2008/56/EC; MSFD), which involve estimating common bottlenose dolphin
152 distribution. We considered two large-scale monitoring programs about bottlenose dolphins.
153 We divided the study area in 4,356 contiguous pixel/grid-cells creating a 5'x5' Mardsen grid
154 (WGS 84) that we used for all the occupancy models we considered. We used data from at-sea
155 surveys over 21,464 km of the French continental shelf (456 grid-cells sampled, 10.46% of the
156 total number of grid-cells). Observers performed monitoring aboard small sailing and motor
157 boats to locate and photo-identify bottlenose dolphins all year long between 2013 and 2015
158 (Labach et al. 2019). At-sea surveys detected 129 distinct bottlenose dolphin groups located in
159 89 different grid-cells. At-sea surveys did not include planned repeated visits, some grid-cells
160 have been visited once, and others have been visited 50 times. Then, using repeated-visits
161 occupancy models to analyze the at-sea monitoring requires considering only the grid-cells
162 sampled multiple times and hence to drop the data collected in grid-cells sampled only once.
163 Single-visit models enable us to include all data, even data collected in grid-cells that were
164 surveyed only once, which make at-sea a relevant candidate for single-visit model
165 implementation. Besides, we considered data collected during aerial line-transects covering
166 24,624 km of the French Exclusive Economic Zone (EEZ), targeting marine megafauna, and
167 following a distance sampling protocol. The survey sampled 1336 grid-cells (i.e. 30.67% of the
168 total number of grid-cells). Aerial surveys produced 130 distinct bottlenose dolphin detections

located in 87 grid-cells. Sampling effort for aerial surveys was homogeneous over the study area with three or four replicates per line-transect between November 2011 and August 2012 (Laran et al., 2017). Because we used occupancy models, we only considered detection/no-detection data, which lead to a binary 0/1 dataset. Hence, multiple sightings detected in the same groups were coded as 1. Thus, we obtain the two aerial and at-sea detection/no-detection datasets that we analyzed with occupancy models. An important assumption of single-season occupancy models is that the latent ecological state of a grid-cell (the z_i's) remains unchanged between the repeated visits (MacKenzie 2006). When monitoring highly mobile species, such as cetaceans, the closure assumption is likely to be violated because individuals can move into and out of the sampling grid-cell. The size of the grid-cells is much lower than dolphins' range of activity. If individuals' movement in and out of the sampling units is random, then the occupancy estimator is unbiased (Kendall et al. 2013). However, it is unlikely the case for bottlenose dolphins because their use of space is driven by ecological and environmental factors, and occupied locations are used only temporarily by individuals (MacKenzie 2006; Neilson et al. 2018). Closure assumption is crucial to the interpretation of occupancy model's parameters. In cases where this assumption is known to be violated, the parameter is usually interpreted as the probability that a location is used by the species as opposed to probability of species presence. In this situation, the occupancy estimator Ψ_i represents the probability that grid-cell i is *used* by the target species (Kendall et al. 2013), being interpreted as *space-use* by bottlenose dolphins. Occupancy and *space-use* refer to distinct ecological concepts. Occupancy describes the species home range that can be defined as the geographic range of occurrence, while space-use refers to the usage made by individuals of the different components of the home range (e.g. feeding locations, migratory routes, Johnson 1980). Then, both single-visit and repeated-visits occupancy models infer the probability that a particular grid-cell is used by the species. The detection probability now accounts for both the probability of detecting the species

given that the species is available for sampling, and the probability that the species is using the grid-cell during the sampling, reflecting that the species might occupy the grid-cell but not during the sampling occasion (MacKenzie 2006). As stated above, single-visit occupancy relaxed the closure assumption because the inference of the detection probability does not require site closure between the repeated visits. However, the interpretation of the occupancy parameter is still space-use in the case of bottlenose dolphins because our single-visit data is collected during multiple years and dolphin are expected to move in and out the sampling unit area during the sampling period.

Because at-sea and aerial surveys were performed during different years, we considered them as independent. In 2018, recent Mediterranean scale aerial monitoring program sampled French Mediterranean following the same line-transect protocol as the aerial dataset we analyzed (ACCOBAMS Survey Initative, 2018). Preliminary and unpublished results from the 2018 program estimated similar common bottlenose dolphin distribution to that of 2011-2012. Then, we assumed that space-use remained unchanged during the monitoring period (i.e. 2011 to 2015). Besides, we neglected the seasonal variation in the bottlenose dolphin space-use in this case study. Concerning the ecological process, we used two environmental covariates to estimate the space-use of bottlenose dolphins: i) bathymetry, which is expected to have a positive effect on bottlenose dolphins' occurrence (Bearzi et al. 2009, Labach et al. 2019), and ii) sea surface temperature (SST, AQUA MODIS | NASA 2019, https://neo.sci.gsfc.nasa.gov/), which is locally related to dolphins' prey abundance and hence expected to affect local distribution of bottlenose dolphins (Bearzi et al. 2009). We extracted average SST between 2011 and 2015 value in each grid-cell, making SST a cell-specific covariate. Similarly, bathymetry had a single value per grid-cell. We checked for correlation between the two covariates and the Pearson coefficient was < 0.3. Then, we modelled Ψ as a function of bathymetry, SST, and the interaction between bathymetry and SST on a logit scale:

$$\text{logit}(\Psi_i) = \beta_0 + \beta_1 \text{ bathymetry}_i + \beta_2 \text{ SST}_i + \beta_3 \, bathymetry_i \, SST_i$$

Regarding the observation process, we calculated the transect length (in km) prospected by each monitoring protocol within each grid-cell during a time period. Sampling effort was therefore a grid-cell-specific and time-specific covariate; Seff_A refers to the sampling effort of the aerial monitoring program while Seff_S refers to the sampling effort of the at-sea monitoring program. We modeled monitoring-specific detection probabilities as functions of the relevant sampling effort:

$$\text{logit}(p_a) = \alpha_0 + \alpha_1 \log(\text{Seff}_A)$$

$$\text{logit}(p_s) = \alpha_0 + \alpha_1 \log(\text{Seff}_S).$$

Regarding the repeated-visit occupancy models, we divided the detection/non-detection datasets into four sampling occasions ($J = 4$): winter (January, February, March), spring (April, May, June), summer (July, August, September), autumn (October, November, December). For the single-visit occupancy models, we considered the entire monitoring program in a single occasion. For example, let us assume that the detection history at site i is $y_i = \{0,1,1,0\}$ in repeated-visit occupancy, i.e. the species was detected at sampling occasions $j = 2$ and $j = 3$, and went undetected at $j = 1$, and $j = 4$, then for single-visit occupancy we have $y_i = \{1\}$. In addition, the single-visit sampling effort in a grid-cell was equal to the sum of the sampling effort over the 4 sampling occasions of the repeated-visit occupancy model.

Performances of integrated models

To assess the added value of combining aerial and at-sea datasets into integrated single-visit occupancy models, we analyzed 3 datasets: i) the aerial dataset, ii) the at-sea dataset, and iii) the two datasets together into an integrated occupancy model. For each of these datasets, we applied repeated-visit and single-visit occupancy models. Besides the case study, we also carried out a simulation study to test for the performances of integrated occupancy models (Appendix S2). In Appendix S5, we go through a worked example of the likelihood functions

for single-visit, repeated-visit, integrated repeated-visit, and integrated single-visit occupancy models. In Appendix S4, we listed the modeling assumptions required to run the different occupancy models.

Bayesian implementation

We ran all models with three Markov Chain Monte Carlo chains with 100,000 iterations each in JAGS (Plummer and others 2003) called from R (R Core Team, v 3.2.5 2019) using the *r2jags* package (Su and Yajima 2015). We checked for convergence calculating the *R-hat* parameter (Gelman et al. 2013) and reported posterior means and 95% credible intervals (CI) for each regression coefficient of covariates affecting space-use probability (Fig. 1). Hereafter, we considered *effect size* of a covariate as the estimate of its regression coefficient. We discussed the effect of a covariate whenever the 95% CI of its associated parameter did not include 0. From covariates' effect size, we calculated the predicted space-use by bottlenose dolphins (i.e. Ψ, Fig. 2). We reported maps of standard deviation of Ψ (Fig. 2B). On the maps, we displayed mean and standard deviation of Ψ for coastal and pelagic seas according to a 500m deep boundary that corresponds to the separation of continental shelf from the abysses. Data and codes are available on Data S1, and on GitHub at https://github.com/valentinlauret/IntegratedSingleVisitOccupancy.

Results

All models produced similar predictions of space used by bottlenose dolphins (Fig. 2). The 95% CI of SST, and of the interaction between SST and bathymetry included 0 in all models (Fig. 1). The probability of space-use increased with decreasing bathymetry for all models (Fig. 1). Bathymetry ranges from altitude of 0 m to -3,488 m deep, hence a positive influence of bathymetry referred to a preference for a high seafloor (e.g. 0-200m depth). Overall, maps showed greater probabilities of space-use on the continental shelf (mean Ψ = 0.76 SD ± 0.17) than on the high seas (mean Ψ = 0.40 SD ± 0.15), although magnitudes of Ψ

269 were different between models (Fig. 2). Bathymetry posterior means were highest for at-sea
270 occupancy (although the 95% CI of effect size included 0), which resulted in models using only
271 at-sea survey data predicting the highest contrast between the continental shelf and the high-
272 seas. Bathymetry effect size was the lowest for aerial occupancy while maps from integrated
273 occupancy models displayed moderate contrast of space-use between shelf and pelagic waters
274 (Fig. 2). Single-visit occupancy models exhibited similar covariates estimates to those of
275 repeated-visit occupancy models (Fig. 1). For aerial occupancy, we noticed similar space-use
276 prediction between single- and repeated-visit (Fig. 2A). For at-sea, predicted space-use
277 probabilities were different between single-visit and repeated-visit occupancy models (Fig. 2).
278 When considering the covariates' effect size (Fig. 1), the widths of the 95% CI were not
279 smaller for integrated occupancy than for occupancy models using datasets in isolation.
280 However, when looking at the standard deviation of the predicted probability of space-use,
281 integrated occupancy models had a better precision than aerial or at-sea occupancy models
282 separately, (Fig. 2B). The use of integrated single-visit occupancy models also improved
283 precision in predicted space-use compared to single-visit occupancy built from aerial and at-
284 sea datasets separately (Fig.2B). Inspecting the simulation results, we found that 1) integrated
285 occupancy models produced more precise estimates of covariates effect size than occupancy
286 models fitted to a single dataset (Appendix S2), and 2) single-visit occupancy models produced
287 similar results to repeated-visit occupancy models (Appendix S1).

288 **Discussion**

289 **Integrated single-visit occupancy models provide reliable ecological inference**

290 Ecological inference from integrated occupancy models lied within the range of the
291 estimates obtained with each dataset separately (Fig. 1). Across all occupancy models, the
292 effects of environmental covariates were similar and consistent with previous studies.
293 Bottlenose dolphins were more likely to use shallower seas (Bearzi et al. 2009, Labach et al.

294 2019), and depth had a stronger effect than SST on the use of space by bottlenose dolphins
295 (Torres et al. 2008). However, we found variations among models in the estimation of the
296 probability of space-use by dolphins (Fig. 1). In particular, at-sea occupancy models predicted
297 that dolphins make little use of the pelagic seas compared to the continental shelf, while aerial
298 occupancy models predict more homogeneous space-use between coasts and pelagic seas.
299 Aerial surveys detected several dolphin groups in the high depths while at-sea surveys detected
300 none. Detecting offshores groups tempered the preference for low-depth seafloors in aerial
301 occupancy models (Appendix S6). Besides, we recommend caution in interpreting predicted
302 maps of space-use as predicted space-use was sensitive to the mean value of covariate effect
303 size. Therefore, depth being the only covariate that affect space-use probability, maps of
304 predicted space-use were mostly driven by bathymetry effect size, and did not account for
305 precision associated with space-use prediction. Because depth posterior mean was similar
306 between occupancy models, differences between predicted space-use maps do not provide a
307 relevant illustration to compare occupancy models performances, nor they reflect the
308 uncertainty associated with the occupancy models' estimates. To study the benefits of single-
309 visit and integrated occupancy models to accommodate existing ecological datasets, we
310 emphasize standard deviation maps and the credible intervals of covariates effect size (Fig. 1-
311 2B). Integrated occupancy models had a better precision in space-use than models using aerial
312 or at-sea surveys separately (Fig. 2). This result was supported by our simulation study which
313 demonstrates the better performance of integrated occupancy models at estimating covariate
314 effect size compared to occupancy models from a single dataset (Appendix S2). Single-visit
315 occupancy models gave similar estimates to those obtained with repeated-visit occupancy
316 models, although repeated-visit occupancy models exhibited better precision (Fig. 1-2B), as
317 well as in our simulations (Appendix S1). In the bottlenose dolphins case study, we considered
318 two existing monitoring programs that were not initially designed for occupancy modeling. In

319 the at-sea monitoring, repeated line-transects were not implemented, nor the high depths were
320 sampled, which made at-sea occupancy unlikely to exhibit precise estimates at our spatial
321 extent. The two datasets exhibit complementary characteristics. While aerial surveys covered a
322 larger spatial scale than at-sea surveys, at-sea surveys exhibited a better detection rate.
323 Detection probability was greater for at-sea surveys ($p = 0.18$ SD $\pm$ 0.04) than for aerial surveys
324 ($p = 0.10$ SD $\pm$ 0.03). Regarding the aerial dataset, the number of occurrences was low despite
325 the important coverage of the monitoring design (i.e. bottlenose dolphins were detected in 6.5%
326 of sampled grid-cells), which might hinder the implementation of single-visit occupancy when
327 the number of occurrences is less than 10% of the sampling units (Peach et al. 2017). However,
328 the at-sea dataset had occurrences in 19.5% of sampled units. Using integrated occupancy
329 models enables to combine low-frequency occurrence data like the aerial dataset with another
330 dataset to increase the amount of information about the ecological state process and helps
331 mitigating the issue of low number of occurrences.

332 **Ecological implications and perspectives**

333 Overall, we illustrate that: i) Integrating datasets into occupancy models improves the
334 precision of space-use estimates, and ii) Single-visit occupancy models can reliably
335 accommodate the lack of repeated visits that occurs frequently. Integrated occupancy models
336 produced more reliable estimates than occupancy models using datasets in isolation in both the
337 bottlenose dolphin data analyzes and the simulations. Our finding on the bottlenose dolphins
338 case study is a good illustration of the well-known benefit of combining datasets into integrated
339 species distribution models to increase precision in ecological inference (Fletcher et al. 2019).
340 Some advanced developments of occupancy models allow combining datasets to estimate
341 occupancy parameters at multiple spatial scales (Nichols et al. 2008, Pavlacky et al. 2012).
342 Besides, integrated occupancy modeling has also been used to evaluate ecological monitoring
343 programs prior to their implementation (e.g., comparing capabilities of different detection

344 devices, Otto & Roloff 2011; Haynes et al. 2013). Here, we emphasize the benefit of
345 considering integrated methods combined with single-visit occupancy modeling after data
346 collection. When the species of interest either occurs over a large spatial scale or is a highly
347 mobile species (such as bottlenose dolphins), considering multiple sampling methods is
348 effective to monitor the entire population making the most of each device (Zipkin and Saunders
349 2018). In particular, integrating a large volume of data, such as those that can be leveraged
350 through citizen-science programs or with dedicated NGOs over the years can make the most of
351 ecological monitoring programs for the furthering of many applied situations (Zipkin et al.
352 2019). However, caution should be taken as integrating data is not always beneficial and
353 requires additional modelling assumptions according to the particularity of each dataset ot
354 include (Dupont et al., 2019; Farr et al., 2020; Fletcher et al., 2019; Simmonds et al., 2020).
355 Although repeated-visit occupancy models remain statistically more precise, there are benefits
356 in using single-visit occupancy models. The ability of single-visit occupancy to relax the
357 closure assumption is appealing, because this assumption is often incompatible with the
358 behavior of mobile species and for numerous monitoring programs of animal populations (Rota
359 et al. 2009, Issaris et al. 2012, Sólymos and Lele 2016, Lele et al. 2012, Kendall et al. 2013).
360 In this study, the closure assumption is unlikely to be valid for bottlenose dolphins over the
361 time span of the two monitoring programs, because dolphins obviously would not remain into
362 the same grid-cell. Besides, when financial or logistical costs are limited, implementing a
363 single-visit monitoring design could provide robust ecological inference while explicitly
364 accounting for imperfect species detection (Lele et al. 2012, Dénes et al. 2017). Overall,
365 increasing quantity and types of biodiversity data are becoming available (Isaac et al. 2019).
366 Numerous monitoring programs do not rely on protocols implementing repeated visits like, e.g.,
367 historical monitoring programs, or citizen science programs (Tingley and Beissinger 2009,
368 Zipkin and Saunders 2018). Then, using single-visit occupancy models helps making efficient

use of available data, which is of great interest in many ecological applications (Nichols and Williams 2006, Sólymos and Lele 2016). In this context, Miller et al. (2019) encouraged further developments of methods mixing standardized and non-standardized datasets. To illustrate, we built an integrated occupancy model mixing repeated-visit occupancy models for aerial surveys and single-visit occupancy models for at-sea surveys (Appendix S3). One could also extend integrated occupancy models to more than two datasets. However, caution should be taken when integrating datasets, as combining different sources of information does not always outperform the analysis of single datasets in isolation (Lele and Allen 2006, Simmonds et al. 2020). The flexibility of occupancy models provided a relevant framework to combine monitoring programs and to accommodate different types of data collection. Integrated and single-visit occupancy models contribute to widen the scope of possibilities. We emphasize the usefulness of both integrated and single-visit approaches to deal with existing datasets. We anticipate that their combination into integrated single-visit approaches will be of most interest for many parties in ecological research.

Acknowledgements

The French Ministry in charge of the environment (Ministère de la Transition Energetique et Solidaire) and the French Office for Biodiversity (OFB) funded the project SAMM. The PELAGIS observatory, with the help of the OFB, designed, coordinated and conducted the survey. We are grateful to all financial partners of the GDEGeM project. We warmly thank technical and scientific participants of GDEGeM. We also thank three anonymous reviewers for their insightful comments and suggestions.

Literature cited

ACCOBAMS Survey Initiative. 2018. ASI 2018 Summer Survey | Accobams. https://accobams.org/main-activites/accobams-survey-initiative-2/asi-preliminary-results/.

Bearzi, G., C. M. Fortuna, and R. R. Reeves. 2009. Ecology and conservation of common bottlenose dolphins *Tursiops truncatus* in the Mediterranean Sea. Mammal Review 39:92–123.

Dénes, F. V., P. Sólymos, S. Lele, L. F. Silveira, and S. R. Beissinger. 2017. Biome-scale signatures of land-use change on raptor abundance: insights from single-visit detection-based models. Journal of Applied Ecology 54:1268–1278.

Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) (Text with EEA relevance). 2008. . Page 164.

Fletcher, R. J., T. J. Hefley, E. P. Robertson, B. Zuckerberg, R. A. McCleery, and R. M. Dorazio. 2019. A practical guide for combining data to model species distributions. Ecology:e02710.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian data analysis. Chapman and Hall/CRC.

Isaac, N. J. B., M. A. Jarzyna, P. Keil, L. I. Dambly, P. H. Boersch-Supan, E. Browning, S. N. Freeman, N. Golding, G. Guillera-Arroita, P. A. Henrys, S. Jarvis, J. Lahoz-Monfort, J. Pagel, O. L. Pescott, R. Schmucki, E. G. Simmonds, and R. B. O'Hara. 2019. Data Integration for Large-Scale Models of Species Distributions. Trends in Ecology & Evolution.

Issaris, Y., S. Katsanevakis, M. Salomidi, K. Tsiamis, N. Katsiaras, and G. Verriopoulos. 2012. Occupancy estimation of marine species: dealing with imperfect detectability. Marine Ecology Progress Series 453:95–106.

Johnson, D. H. 1980. The Comparison of Usage and Availability Measurements for Evaluating Resource Preference. Ecology 61:65–71.

Kendall, W. L., J. E. Hines, J. D. Nichols, and E. H. C. Grant. 2013. Relaxing the closure

assumption in occupancy models: staggered arrival and departure times. Ecology 94:610–617.

Knape, J., and F. Korner-Nievergelt. 2015. Estimates from non-replicated population surveys rely on critical assumptions. Methods in Ecology and Evolution 6:298–306.

Labach, H., C. Azzinari, M. Barbier, C. Cesarini, B. Daniel, L. David, F. Dhermain, N. Di-Méglio, B. Guichard, J. Jourdan, N. Robert, M. Roul, N. Tomasi, and O. Gimenez. 2019. Distribution and abundance of bottlenose dolphin over the French Mediterranean continental shelf. bioRxiv.

Laran, S., E. Pettex, M. Authier, A. Blanck, L. David, G. Dorémus, H. Falchetto, P. Monestiez, O. Van Canneyt, and V. Ridoux. 2017. Seasonal distribution and abundance of cetaceans within French waters- Part I: The North-Western Mediterranean, including the Pelagos sanctuary. Deep Sea Research Part II: Topical Studies in Oceanography 141:20–30.

Lele, S. R., and K. L. Allen. 2006. On using expert opinion in ecological analyses: a frequentist approach. Environmetrics 17:683–704.

Lele, S. R., M. Moreno, and E. Bayne. 2012. Dealing with detection error in site occupancy surveys: what can we do with a single survey? Journal of Plant Ecology 5:22–31.

Lindenmayer, D. B., and G. E. Likens. 2010. The science and application of ecological monitoring. Biological Conservation 143:1317–1328.

MacKenzie, D. I., editor. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species. Elsevier, Amsterdam ; Boston.

Mackenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one 83:8.

Miller, D. A. W., K. Pacifici, J. S. Sanderlin, and B. J. Reich. 2019. The recent past and promising future for data integration methods to estimate species' distributions. Methods in Ecology and Evolution 10:22–37.

Neilson, E. W., T. Avgar, A. C. Burton, K. Broadley, and S. Boutin. 2018. Animal movement affects interpretation of occupancy models from camera-trap surveys of unmarked animals. Ecosphere 9:e02092.

Nichols, J. D., L. L. Bailey, A. F. O'Connell Jr., N. W. Talancy, E. H. Campbell Grant, A. T. Gilbert, E. M. Annand, T. P. Husband, and J. E. Hines. 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. Journal of Applied Ecology 45:1321–1329.

Nichols, and Williams. 2006. Monitoring for conservation. Trends in Ecology & Evolution 21:668–673.

Pavlacky, D. C., J. A. Blakesley, G. C. White, D. J. Hanni, and P. M. Lukacs. 2012. Hierarchical multi-scale occupancy estimation for monitoring wildlife populations. The Journal of Wildlife Management 76:154–162.

Peach, M. A., J. B. Cohen, and J. L. Frair. 2017. Single-visit dynamic occupancy models: an approach to account for imperfect detection with Atlas data. Journal of Applied Ecology 54:2033–2042.

Plummer, M. and others. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Page 125 Proceedings of the 3rd international workshop on distributed statistical computing. Vienna.

R Core Team, v 3.2.5. 2015. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. https://www.r-project.org/.

Rota, C. T., R. J. Fletcher Jr, R. M. Dorazio, and M. G. Betts. 2009. Occupancy estimation and the closure assumption. Journal of Applied Ecology.

Royle, J. A., and M. Kéry. 2007. A Bayesian state-space formulation of dynamic occupancy models. Ecology 88:1813–1823.

Simmonds, E. G., S. G. Jarvis, P. A. Henrys, N. J. B. Isaac, and R. B. O'Hara. 2020. Is more data always better? A simulation study of benefits and limitations of integrated distribution models. Ecography:ecog.05146.

Sólymos, P., and S. R. Lele. 2016. Revisiting resource selection probability functions and single-visit methods: clarification and extensions. Methods in Ecology and Evolution 7:196–205.

Su, Y.-S., and M. Yajima. 2015. R2jags: Using R to Run "JAGS."

Tingley, M. W., and S. R. Beissinger. 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. Trends in Ecology & Evolution 24:625–633.

Torres, L. G., A. J. Read, and P. Halpin. 2008. Fine-scale habitat modeling of a top marine predator: do prey data improve predictive capacity. Ecological Applications 18:1702–1717.

Zipkin, E. F., B. D. Inouye, and S. R. Beissinger. 2019. Innovations in data integration for modeling populations. Ecology:e02713.

Zipkin, E. F., and S. P. Saunders. 2018. Synthesizing multiple data types for biological conservation using integrated population models. Biological Conservation 217:240–250.

Zipkin, E. F., E. R. Zylstra, A. D. Wright, S. P. Saunders, A. O. Finley, M. C. Dietze, S. Itter, and M. W. Tingley. 2021. Addressing data integration challenges to link ecological processes across scales. Frontiers in Ecology and the Environment:9.

Figure captions

Figure 1: Effect size of bathymetry, sea surface temperature (SST), and interaction between SST and bathymetry on the probability Ψ that a site is used by Bottlenose dolphins (*Tursiops truncatus*). The posterior mean is provided with the associated 95% credible interval. "SV" refers to *single-visit* occupancy models, "RV" to *repeated visit* occupancy models, and "IOM" stands for *integrated* occupancy models, in which aerial surveys and at-sea surveys are combined. Estimates are given on the logit scale.

Figure 2: A. Probability of predicted space-use by Bottlenose dolphins (*Tursiops truncatus*) over the NW Mediterranean Sea. Using the posterior mean of covariates effect size, we estimated the probability that a grid-cell was used by bottlenose dolphins. For each occupancy model, we added the mean space-use probability (Ψ) for coasts (bathymetry < 500 m) and pelagic seas (bathymetry > 500 m)

B. Standard deviation of predicted space-use. Using the posterior standard deviation of covariates effect size, we estimated the standard deviation associated with the space-use probability. For each occupancy model, we added the mean standard-deviation (sd) associated with Ψ for coasts (bathymetry < 500 m) and pelagic seas (bathymetry > 500 m).

"IOM" stands for *integrated* occupancy models, in which aerial surveys and at-sea surveys are combined. Repeated-visit occupancy maps refer to occupancy models with 4 sampling occasions. Single-visit maps refer to occupancy models considering 1 sampling occasion.

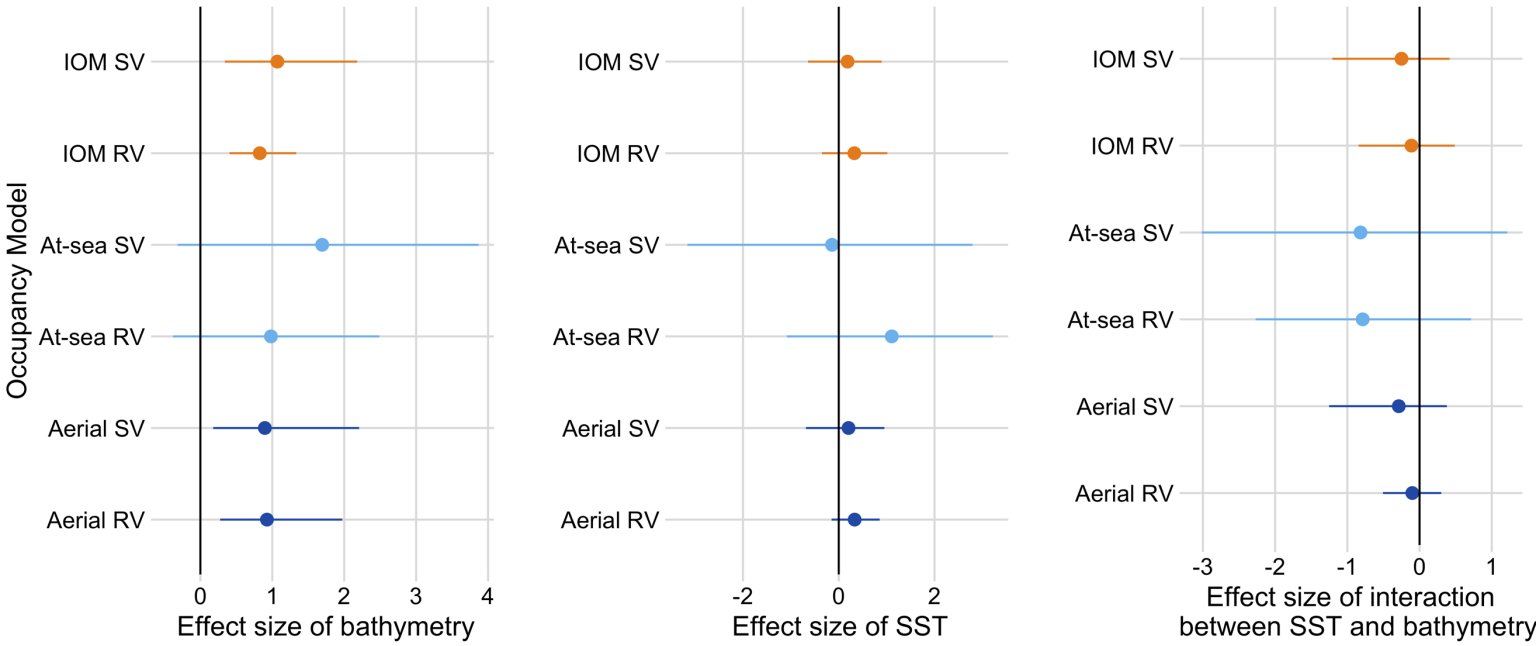
Occupancy Model
IOM SV
IOM RV
At-sea SV
At-sea RV
Aerial SV
Aerial RV
0
1
2
3
4
Effect size of bathymetry
-2
0
2
Effect size of SST
-3
-2
-1
0
1
Effect size of interaction
between SST and bathymetry

Figure 1

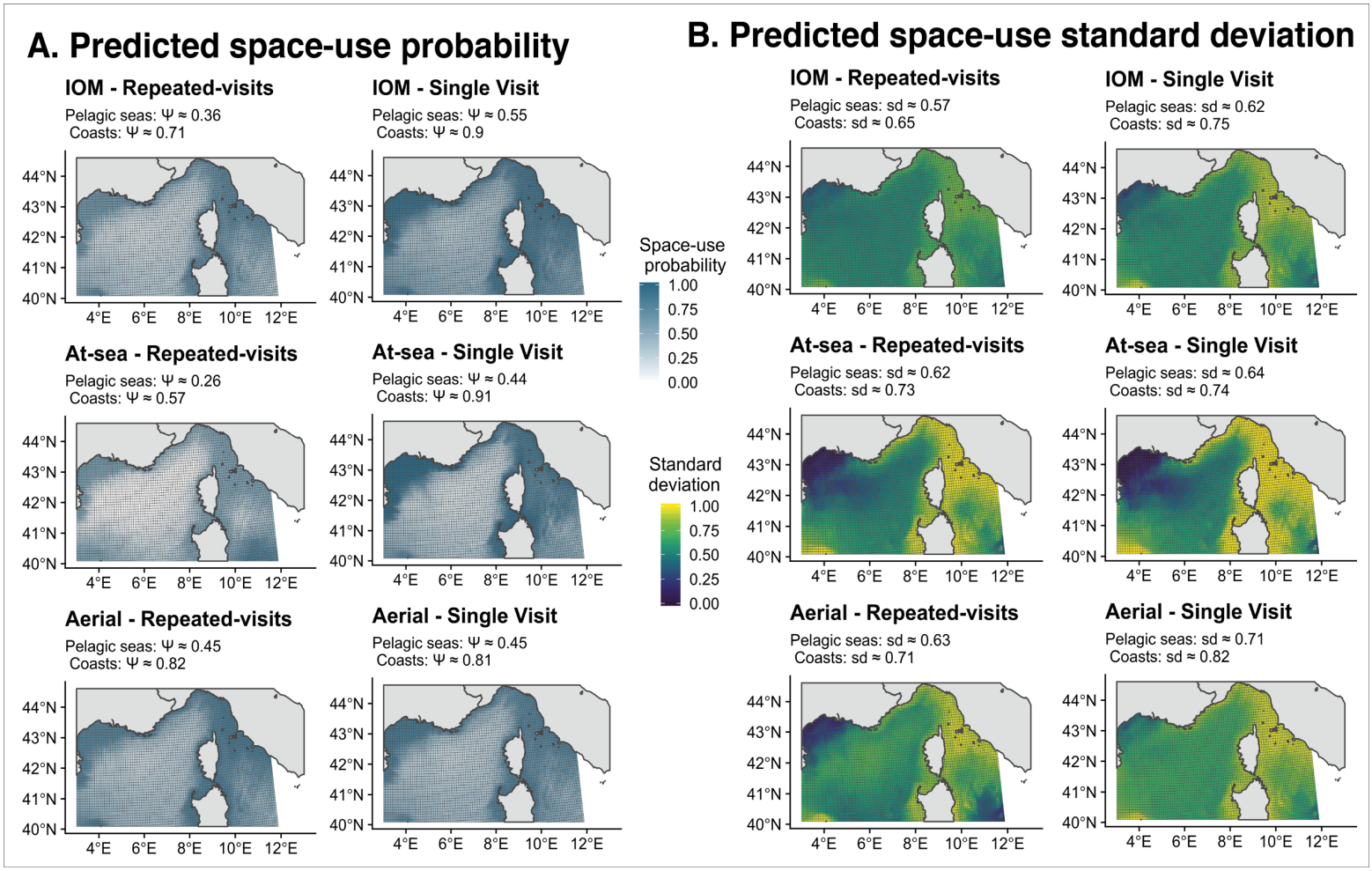
A. Predicted space-use probability
IOM - Repeated-visits
Pelagic seas: Ψ ≈ 0.36
Coasts: Ψ ≈ 0.71
IOM - Single Visit
Pelagic seas: Ψ ≈ 0.55
Coasts: Ψ ≈ 0.9
At-sea - Repeated-visits
Pelagic seas: Ψ ≈ 0.26
Coasts: Ψ ≈ 0.57
At-sea - Single Visit
Pelagic seas: Ψ ≈ 0.44
Coasts: Ψ ≈ 0.91
Aerial - Repeated-visits
Pelagic seas: Ψ ≈ 0.45
Coasts: Ψ ≈ 0.82
Aerial - Single Visit
Pelagic seas: Ψ ≈ 0.45
Coasts: Ψ ≈ 0.81
Space-use probability
1.00
0.75
0.50
0.25
0.00
Standard deviation
1.00
0.75
0.50
0.25
0.00
B. Predicted space-use standard deviation
IOM - Repeated-visits
Pelagic seas: sd ≈ 0.57
Coasts: sd ≈ 0.65
IOM - Single Visit
Pelagic seas: sd ≈ 0.62
Coasts: sd ≈ 0.75
At-sea - Repeated-visits
Pelagic seas: sd ≈ 0.62
Coasts: sd ≈ 0.73
At-sea - Single Visit
Pelagic seas: sd ≈ 0.64
Coasts: sd ≈ 0.74
Aerial - Repeated-visits
Pelagic seas: sd ≈ 0.63
Coasts: sd ≈ 0.71
Aerial - Single Visit
Pelagic seas: sd ≈ 0.71
Coasts: sd ≈ 0.82
44°N
43°N
42°N
41°N
40°N
4°E
6°E
8°E
10°E
12°E

Figure 2